基于权值分布的多模型分类算法研究*

蒋梦莹[1,2]，林小竹[1]，柯　岩[1,2]，魏战红[1]

(1. 北京石油化工学院 信息工程学院，北京 102617; 2. 北京化工大学 信息科学与技术学院，北京 100029)

摘　要：为了提高卷积神经网络对图像分类的正确率，对网络结构进行研究，提出了多模型融合卷积神经网络。通过提取单个模型的输出特征向量，进行融合后得到新的输出特征向量，再搭建单层分类器进行图像分类，提高分类准确率。将单个模型与多模型融合的分类准确率进行比较，多模型融合卷积神经网络的分类准确率有所提高。并分析了卷积神经网络最后一层全连接层的权值分布，发现同一模型在不同数据集上权值分布曲线相似，分类效果好的网络模型其权值分布曲线更平缓。

关键词：卷积神经网络；多模型融合；特征向量；特征提取；权值分布

中图分类号：TP391.4　　doi: 10.19734/j.issn.1001-3695.2018.05.0506

Research on multi-model classification algorithm based on weight distribution

Jiang Mengying[1, 2], Lin Xiaozhu[1], Ke Yan[1, 2], Wei Zhanhong[1]

(1. *College of Information Engineering*, *Beijing Institute of Petrochemical Technology*, *Beijing* 102617, *China*; 2. *College of Information Science & Technology*, *Beijing University of Chemical Technology*, *Beijing* 100029, *China*)

Abstract: To improve the correct rate of image classification by convolutional neural network, a multi-model fusion convolutional neural network is proposed after research on the network structure. By extracting the output feature vectors of a single model and then fusion them, the new output feature vectors are obtained, and then a single classifier is set up to classify the images, and the accuracy of the classification is improved. The classification accuracy of single model compare with multi-model fusion, the accuracy of classification of multi-model fusion convolutional neural network is improved. The weight distribution of the last layer of the convolutional neural network is analyzed, and it is found that the weight distribution curve of the same model on different data sets is similar and the weight distribution curve of the network model with better classification effect is more gentle.

Key words: convolutional neural network; multi-model fusion; feature vector; feature extraction; weight distribution

0　引言

图像的特征提取和分类一直以来都是计算机视觉领域一个基础且重要的研究方向。深度学习是近十余年来人工智能领域取得的重要突破，其卷积神经网络在图像分类上具有显著效果。卷积神经网络（convolutional neural network，CNN）能从原始输入图片中挖掘抽象的特征信息，得到的特征表示具有良好的泛化能力。1998 年 LeCun 提出了 LeNet-5 网络结构[1]标志着 CNN 的真正面世，但此时的 CNN 只适用于小图片的识别，对于大规模的数据输入，识别效果较差。直至 2012 年，krizhevsky 等人[2]设计的 AlexNet 卷积神经网络结构在 ImageNet 图像竞赛（ImageNet Large Scale Visual Recognition Challenge,LSVRC）中取得冠军，使 CNN 得到越来越多的研究者们的重视。2014 年，Szegedy[3]等人增加了 CNN 的深度，提出了超出 20 层的 CNN 结构 GoogLeNet 模型。Simonyan 等人[4]对 CNN 的深度进行研究，提出了 VGGNet 模型，通过不断增加 3×3 卷积核的卷积层来增加网络的深度。2015 年，He 等人[5]提出深度为 152 层的残差网络在 LSVRC-15 的图像分类竞赛中获得了第一名。随着卷积神经网络结构的不断加深，模型需要海量的数据输入。由于数据集数量有限，训练网络存在梯度弥散和过拟合[6]等问题，导致图片分类的准确率下降。对此，研究者们提出了许多解决办法。在网络结构方面，通过改变网络的激活函数[7]、引入 dropout 层[8]、迁移学习、并行交叉 CNN 模型[9]和本文前期工作——跨连接[10]等方法来提高模型对特征的表达能力，进而提高图片分类的准确率。本文首先提出一种多模型融合的卷积神经网络来进行图像分类，将数据集采用迁移学习[11]在几个现有的卷积神经网络

收稿日期：2018-05-09；修回日期：2018-06-30　　基金项目：国家自然科学基金资助项目（61702040）；北京市自然科学基金资助项目（4174089）

作者简介：蒋梦莹（1994-），女，江西宜春人，硕士研究生，主要研究方向为图像处理与模式识别、深度学习等；林小竹（1964-），男（通信作者），教授，博士，主要研究方向为图像处理与模式识别、深度学习等（linzhu1964@163.com）；柯岩（1994-），男，硕士研究生，主要研究方向为主要研究方向为图像处理与模式识别、深度学习等；魏战红（1985-），女，讲师，博士，主要研究方向为图像处理和图像信息隐藏.

模型中进行训练，并分别提取训练后模型的最后一层全连接层输出特征向量。将几个模型提取出的特征向量进行融合，使用融合后的特征向量进行 Softmax 分类，提高分类准确率，并通过实验验证了方法的可行性。然后对单模型和多模型的最后一层全连接层的权值分布曲线进行分析。

1　多模型融合卷积神经网络

卷积神经网络可以直接输入原始图像，避免了图片复杂的前期预处理，不需要人为地进行特征提取，具有良好的特征表达能力[12]。卷积神经网络主要包括卷积层、下采样层、局部响应归一化（local response normalization，LRN）层[13]、全连接层和 Softmax 分类层，其中卷积层和池化层是卷积神经网络的重要组成部分。卷积神经网络的训练过程主要包括前向传播和反向传播两个过程[14]，主要是学习卷积层的卷积核参数和层间连接权重等网络参数。

1.1　AlexNet 模型和 VGG 模型

1.1.1 AlexNet 模型

AlexNet 模型解决了之前卷积神经网络只能满足小尺寸图片输入的问题，并证明了 CNN 在复杂模型下的有效性。相较于传统的卷积神经网络，AlexNet 模型在网络模型结构、激活函数以及引入 Dropout 层等方面进行了改进。AlexNet 模型增加了网络模型层数，更深的模型结构使其具有更好的特征提取和特征表达能力。局部响应归一化是对特征图相邻区域或者相邻特征图的响应进行归一化操作，起到相邻响应相互抑制的作用。在激活函数方面，摒弃了传统的 sigmoid 激活函数和 tanh 激活函数，采用 ReLU 分段线性激活函数，使得网络能够快速收敛，提供了神经网络的稀疏表达能力并且有效缓解了梯度消失的问题。在全连接层采用 Dropout，即在训练过程中，按照一定的概率将部分神经网络单元随机暂时的从网络中丢弃，增强了模型的泛化能力，改善模型过拟合的问题。

1.1.2 VGGNet 模型

VGGNet 模型是 AlexNet 模型发展而来的网络模型，主要展示出网络的深度是算法优良性能的关键部分。VGGNet 主要在两方面提出了改进：所有的卷积核大小都使用 3×3 尺寸大小；在整个图片和 multi-scale 上训练和测试图片。VGGNet 根据网络层数不同有多种模型，常使用的为 VGG-16 模型，其结构图如图 1 所示。

1.2　多模型融合卷积神经网络

训练卷积神经网络模型时常出现模型过拟合的问题，导致网络模型对图像分类的正确率降低。由于输入图片的数量有限，而网络结构参数多，特征向量在进行分类时全连接层参数容易出现冗余。

对网络模型的全连接层分析发现，改善模型过拟合现象提高分类正确率方法主要分为减少全连接层权值参数和增加输出特征向量两方面。在这基础上，提出一种多模型融合的卷积神经网络模型，对同一训练数据集，提取不同网络模型经过迁移学习后的最后一层全连接层输出特征向量，然后将得到的不同模型的输出特征向量进行融合后，再搭建新的单层分类器进行分类，其模型结构如图 2 所示。

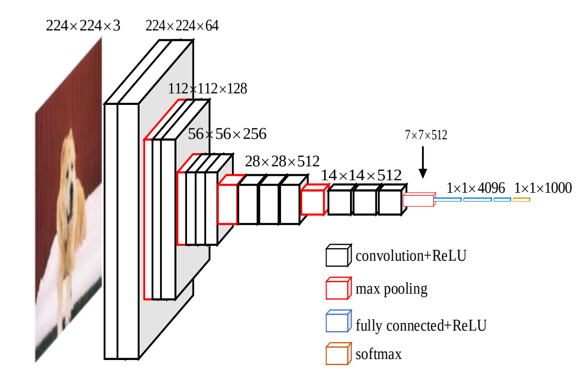


图 1　VGG-16 模型结构图

Fig.16　model

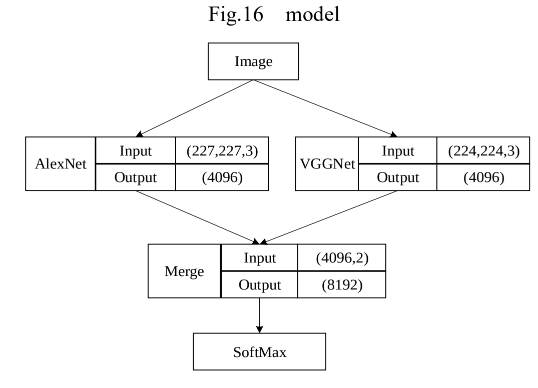


图 2　多模型融合结构图

Fig.2　Multi-model fusion structure diagram

将不同网络模型的输出特征向量进行融合后，增加的输出特征向量参与最后分类器的分类，对分类正确率有一定的提高。同时，在获取不同模型的输出特征向量时，仅采用预训练模型微调网络参数后来得到最后一层全连接层网络的输出特征向量。融合不同模型的输出特征向量后，只需要进行单层分类器训练分类，一定程度的减少了计算量，对训练时的硬件要求降低。

2　CNN 最后一层权值分布

通过对卷积神经网络的结构分析，可以发现卷积神经网络的卷积层和池化层的作用旨在提取图像特征，卷积神经网络后几层全连接层的作用则是将提取的特征变换得到一维的输出特征向量。最后一层全连接的输出特征向量的每个单元可以看做训练样本中含有的特征，经过最后一层全连接层后得到一个一维向量（$1*n$），其中 n 为样本数据集的类别数。网络的最后一层全连接层的作用实际是对特征向量进行分类。对最后一层全连接层进行研究，获取网络模型的最后一层权值参数，并对其权值参数分布进行分析。假设输出特征向量单元数为 4096 个单元，最后一层全连接层过程如图 3 所示。

图中，全连接层的输入为 $\{X_1,X_2,X_3,\ldots,X_{4096}\}$，输出为 $\{a_1,a_2,\ldots,a_n\}$，n 表示分类任务的类别数，w 为全连接层的权值参数。那么，全连接层也可以用一个多维方程组表示，如下式所示:

$$
\begin{aligned}
a_1 &= w_{11} * x_1 + w_{12} * x_2 + \dots + w_{1,4096} * x_{4096} + b_1 \\
a_2 &= w_{21} * x_1 + w_{22} * x_2 + \dots + w_{2,4096} * x_{4096} + b_2 \\
a_3 &= w_{31} * x_1 + w_{32} * x_2 + \dots + w_{3,4096} * x_{4096} + b_3 \\
&\dots \\
a_n &= w_{n1} * x_1 + w_{n2} * x_2 + \dots + w_{n,4096} * x_{4096} + b_n
\end{aligned}
$$

通过公式可知，每一种分类类别都是由 4096 个特征按照一定的权值比重组成的。在进行分类时，通过分析可知，全连接层输入与权值相乘得到的净激活再经过 ReLU 激活函数，净激活小于等于零的值经过 ReLU 激活函数后被置零，只保留大于零的净激活值。那么实际上，在最后一层全连接层进行分类时，只有部分的输出特征决定着分类类别。所以，最后一层全连接层的权值分布情况很大程度上决定着模型分类的效果。

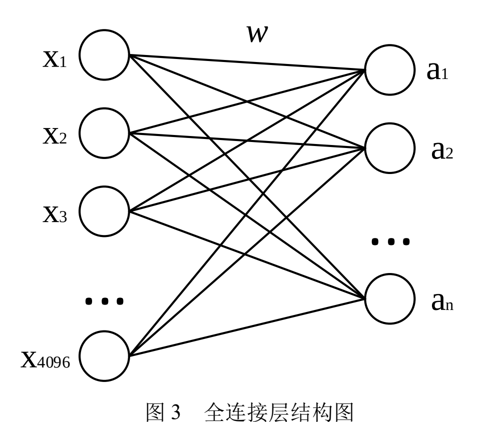


图 3　全连接层结构图

Fig.3　Full connect layer structure diagram

3　实验分析

3.1　实验数据集

本实验分别采用 Caltech-101 数据集[15]和 2017 年百度图像竞赛数据集进行实验。Caltech-101 数据集是由李飞飞等人于 2003 年在加州理工学院创建的数字图像集。Caltech-101 数据集共 9146 张图片，包括 101 类前景图片和一个背景类。每类有 30~800 张图片，大部分图片有 50 张。图 4 为 Caltech-101 数据集的一部分示例图片。

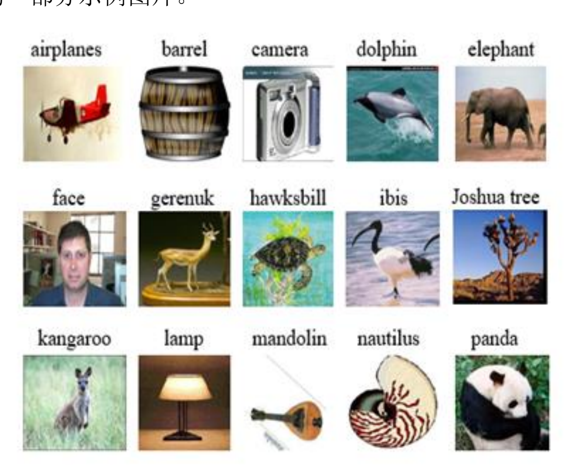


图 4　Caltech-101 数据集的部分图片

Fig.4　Partial pictures of Caltech-101 dataset

2017 百度图像竞赛数据集为 100 类不同品种的狗狗图片，数据集如图 5 所示。其中，训练集数量为 8153 张，测试集数量为 10 624 张，每类有 30~200 张。



图 5　2017 百度图像竞赛数据集部分图片

Fig.5　Partial pictures of 2017 Baidu image competition dataset

3.2　实验结果及分析

3.2.1 多模型融合

本文将 Caltech-101 数据集中每类随机选取出 30 幅图像作为训练样本，剩余部分作为测试样本。将 2017 百度图像竞赛数据集训练集 8153 张图片进行训练，测试集 10 624 张图片进行测试。首先对图像进行预处理，将数据集所有图片缩放为 256*256 像素大小的图像块。由于从 ImageNet 数据集上学习得到的底层滤波器往往描述了各种不同的局部边缘和纹理信息，这些滤波器对一般的图像有较好的普适性。在不同模型提取最后一层全连接层输出向量时，均采用各个模型在 ImageNet 数据集训练过后的参数作为初始权值参数，然后使用输入数据集对网络模型进行微调。使用 CaffeNet 模型进行微调提取输出特征向量时，对输入图像进行处理得到 227*227 像素的图像，然后对所有图像进行减均值处理。CaffeNet 模型与 AlexNet 结构模型基本相似，区别在于 LRN 层的位置不同。AlexNet 结构中，LRN 层是在池化层之前，而 CaffeNet 结构中，LRN 层则在池化层之后。使用 VGG-16 网络模型进行微调提取输出向量时，则截取 224*224 像素的图像输入。初始学习率为 0.0001，每迭代 1000 次学习率减少 10 倍，训练过程总迭代次数为 8000 次。将两个网络模型融合后，得到输出向量为 8192 个单元，再进行单层的 Softmax 分类器分类。

使用 Caltech-101 数据集在 CaffeNet 网络模型提取输出特征向量进行单层分类器训练、在 VGG-16 网络模型提取输出特征向量进行单层分类器训练、将两个模型的输出特征向量融合后进行单层分类器训练，其测试集结果如表 1 所示。

表 1　不同模型在 Caltech-101 数据集上的分类正确率

网络模型	正确率（%）
CaffeNet 模型	79.02
VGGNet 模型	82.25
多模型融合	83.89

将 2017 百度图像竞赛数据集在 CaffeNet 模型提取输出特征向量进行单层分类器训练、在 VGG-16 网络模型提取输出特征向量进行单层分类器训练以及将两个模型的输出特征向量融合后进行单层分类器训练，其测试集结果如表 2 所示。

表 2　不同模型在 2017 百度图像竞赛数据集上的分类正确率

网络模型	正确率（%）
CaffeNet 模型	39.05
VGGNet 模型	42.88
多模型融合	45.67

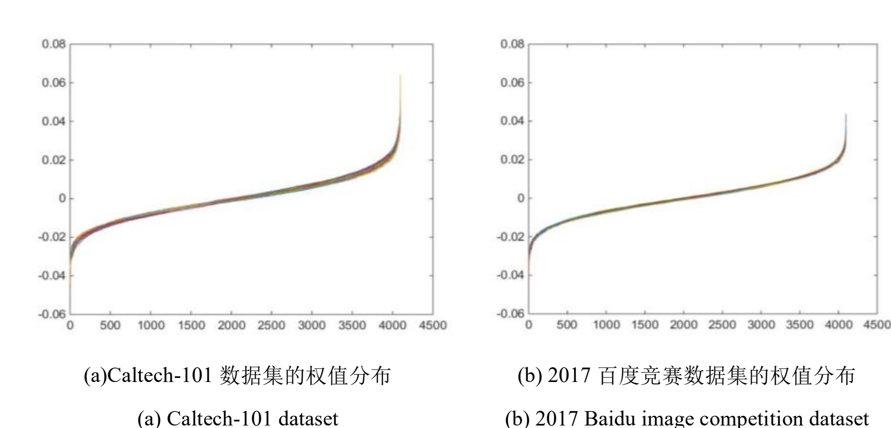
对两个数据集分别进行多个模型训练后，测试结果表明：同一训练样本数据集，VGG-16 模型训练后分类正确率高于 CaffeNet 模型，多模型融合后，分类正确率均高于单个模型训练的分类准确率。

3.2.2 最后一层全连接层权值分析

通过上一节实验可以得到已经训练好的不同卷积神经网络模型，分别提取其最后一层全连接层的权值，将它们按照大小顺序绘制出来。其中，CaffeNet 模型在 Caltech-101 数据集与 2017 百度图像竞赛数据集上训练后最后一层全连接层权值分布如图 6 所示。

图 6 中，横坐标为最后一层全连接层特征向量的维度，一般情况下输出向量的单元数为 4096，横坐标范围为（0，4500）。纵坐标为最后一层全连接层的权值参数值，其范围为（-0.06，0.08）。对比同一模型在不同数据集上的最后一层权值分布曲线，可以得到，同一模型在不同数据集上的权值分布曲线相似。

分别绘制 CaffeNet 模型和 VGG-16 模型在 Caltech-101 数据集上最后一层全连接层权值分布曲线，如图 7 所示。通过分析不同模型在同一数据集上的最后一层权值分布，可以得到，VGG-16 模型相较于 CaffeNet 模型，其最后一层全连接层的权值参数分布更平缓。

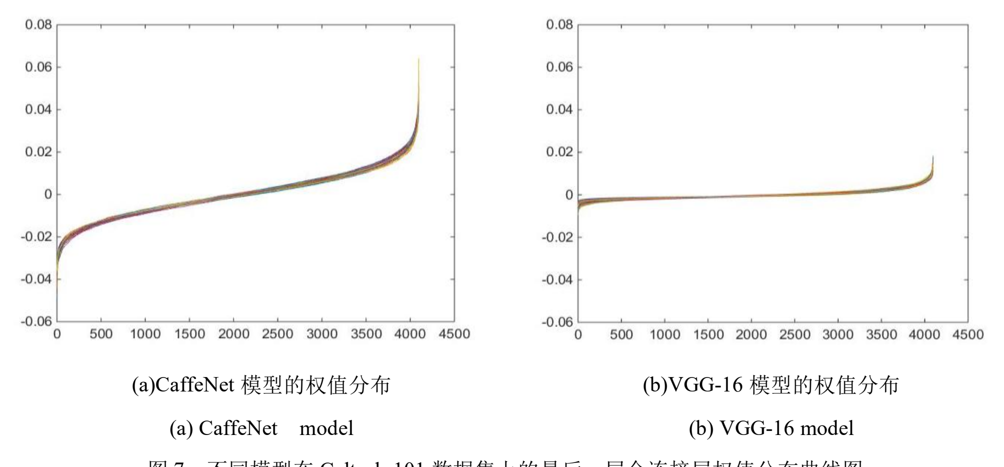


(a)Caltech-101 数据集的权值分布　　(b) 2017 百度竞赛数据集的权值分布

(a) Caltech-101 dataset　　(b) 2017 Baidu image competition dataset

图 6　CaffeNet 模型在不同数据集上的最后一层全连接层权值分布图

Fig.6　The last layer full connected weight distribution on different data sets on CaffeNet model



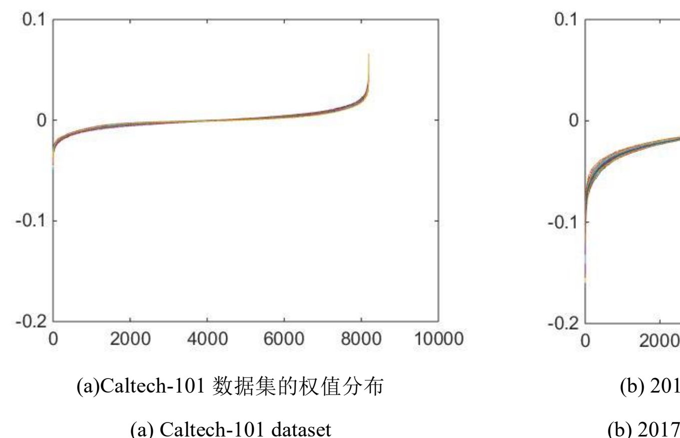
(a)CaffeNet 模型的权值分布　　(b)VGG-16 模型的权值分布

(a) CaffeNet　model　　(b) VGG-16 model

图 7　不同模型在 Caltech-101 数据集上的最后一层全连接层权值分布曲线图

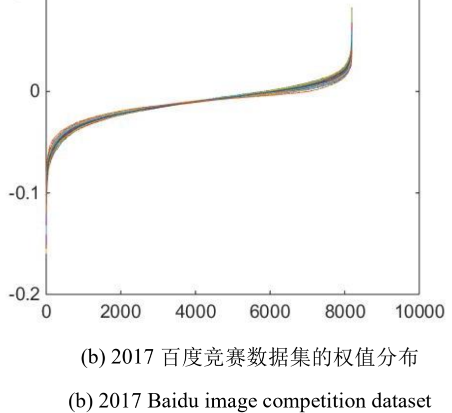
Fig.7　The last layer full connect layer weight distribution of different models on Caltech-101 dataset

对于多模型融合卷积神经网络模型则同样提取其分别在 Caltech-101 数据集和 2017 百度数据集上训练后的最后一层权值分布曲线，如图 8 所示。图中横坐标为最后一层全连接层特征向量的维度，由于多模型融合了两个单模型的输出向量，所以其多模型的特征向量维度为 8192，横坐标范围为（0，10000）。纵坐标为最后一层全连接层的权值参数值，其取值范围为（-0.2，0.1）。由于 2017 年百度图像竞赛数据集相较于 Caltech-101 数据集来说，其数据未经过清洗，图片质量不好，所以分类正确率低。通过对比多模型融合卷积神经网络模型在两个数据集上的权值分布曲线，可以发现多模型融合后权值分布取决于数据集的质量，曲线总体相似，当数据集质量越好，分类曲线越平缓。



(a)Caltech-101 数据集的权值分布

(a) Caltech-101 dataset



(b) 2017 百度竞赛数据集的权值分布

(b) 2017 Baidu image competition dataset

图 8 融合模型在不同数据集上的最后一层全连接层权值分布曲线图

Fig.8 The last layer full link layer weight distribution of fusion models on different datasets

4 结束语

卷积神经网络的特点是能够自动地、隐式地学习特征，不需要人为地定义特征。当有足够多的样本进行训练时，网络可以学习到很好的特征来进行分类。由于网络权值参数过多，训练样本少，网络模型易出现过拟合问题，分类正确率降低。网络模型结构复杂，计算量大，硬件设施要求较高。本文首先提出多模型融合的卷积神经网络模型，其分类准确率较于单个模型有一定的提高。由于多模型融合后，只需进行单层网络训练，对于硬件设施要求也降低，计算速度也得到提高。然后对卷积神经网络的最后一层全连接层进行分析，发现同一模型在不同数据集上最后一层全连接层权值分布曲线相似，不同模型在相同数据集上的最后一层全连接层权值分布曲线有差异。多模型融合后权值分布则与数据集的质量有关，质量越好分类曲线越平缓，而且一致性较好。

参考文献：

[1] Lécun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097-1105.

[3] Szegedy C, Liu Wei, Jia Yangqing, *et al.* Going deeper with convolutions [C]// Computer Vision and Pattern Recognition. 2015: 1-9.

[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014.

[5] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep Residual Learning for Image Recognition [C]// Computer Vision and Pattern Recognition. 2016: 770-778.

[6] Sermanet P, Eigen D, Zhang Xiang, *et al.* OverFeat: Integrated recognition, localization and detection using convolutional networks [J]. Eprint Arxiv, 2013.

[7] Goodfellow I J, Wardefarley D, Mirza M, *et al.* Maxout networks [J]. Computer Science, 2013: 1319-1327.

[8] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15 (1): 1929-1958.

[9] 汤鹏杰，王瀚漓，左凌轩. 并行交叉的深度卷积神经网络模型 [J]. 中国图象图形学报, 2016, 21 (3): 339-347. (Tang Pengji, Wang Hanli, Zuo Lingxuan. Parallel cross deep convolution neural networks model [J]. Journal of Image and Graphics, 2016, 21 (3): 339-347.)

[10] 李勇，林小竹，蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别 [J]. 自动化学报, 2018, 44 (1): 176-182. (Li Yong, Lin Xiaozhu, Jiang Mengying. Facial expression recognition with cross-connect LeNet-5 network [J]. Acta Automatica Sinica, 2018, 44 (1): 176-182)

[11] Pan Sinno Jalin, Yang Qiang. A Survey on Transfer Learning [J]. IEEE Trans on Knowledge & Data Engineering, 2010, 22 (10): 1345-1359.

[12] 常亮，邓小明，周明全，等. 图像理解中的卷积神经网络 [J]. 自动化学报, 2016, 42 (9): 1300-1312. (Chang Liang, Deng Xiaoming, Zhou Mingquan, *et al.* Convolutional neural networks in image understanding [J]. Acta Automatica Sinica, 2016, 42 (9): 1300-1312)

[13] Fleming A D, Philip S, Goatman K A, *et al.* Automated microaneurysm detection using local contrast normalization and local vessel detection [J]. IEEE Trans Med Imaging, 2006, 25 (9): 1223-32.

[14] Bouvrie J. Notes on Convolutional Neural Networks [J]. Neural Nets, 2006.

[15] Borji A, Sihite D N, Itti L. Salient Object Detection: A Benchmark [M]// Computer Vision. Berlin: Springer, 2012: 414-429.